



Improving Early Detection of Learning Disabilities Using Graph Convolutional Networks (GCN) and Multi-Modal Data

Abdumutalliev Abdulkhad Abdusamad ugli^{1*}, Dr.P. Senthil Raja², V. Dhanalakshmi³, M. Gayathri⁴, Zuhra Xudayarova⁵, Mavluda Nurillaeva⁶

¹Turan International University, Namangan, Uzbekistan. E-mail: aabdulaxadturan@gmail.com, <https://orcid.org/0009-0006-1675-0756>

²Assistant Professor, Department of Computer Science and Engineering, K.S. Rangasamy College of Technology, Tiruchengode, India. E-mail: senthilraja@ksrct.ac.in, <https://orcid.org/0000-0002-5843-961X>

³Assistant Professor, Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Tamil Nadu, India. E-mail: dhanalakshmi@maher.ac.in

⁴Assistant Professor, Department of Management Studies, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Tamil Nadu, India. E-mail: gayathrimba@maher.ac.in

⁵Department of Accounting and Statistics, Termez University of Economics and Service, Termez, Uzbekistan. E-mail: zuxra_xudayarova@tues.uz, <https://orcid.org/0009-0000-8767-6975>,

⁶1st Year Master's Degree Student, Linguistics, English Language, Department of English Linguistics, Bukhara State University, Bukhara, Uzbekistan. E-mail: isayevamavluda14@gmail.com, <https://orcid.org/0009-0008-7517-2305>

*Corresponding author: Email: aabdulaxadturan@gmail.com

Abstract

Learning disabilities such as dyslexia, dyscalculia, ADHD, and dysgraphia remain among the most underdiagnosed neurodevelopmental disorders in early childhood, often leading to long-term academic and psychological challenges when not identified at an early stage. The purpose of this study is to design an accurate and interpretable artificial intelligence framework that will facilitate the earlier detection of learning disabilities using an attention-based graph convolutional network with multi-modal data. By utilizing multiple sources of heterogeneous neurocognitive data (e.g., eye-tracking metrics, fMRI, EEG, neuropsychological assessment scores, demographic data) in a unified graph representation, the authors propose to capture complex relational dependencies between the different modalities. The proposed methodology includes multi-stage preprocessing, constructing graphs using connections within the functional and cross-modal data, performing attention-based GCN learning, and performing interpretability analysis using GNNExplainer. The framework was evaluated using a dataset containing 1240 children aged 4-10 years and five diagnostic groups through the use of stratified 5-fold cross-validation. The results of the experiments showed that the overall performance of this method (i.e., MM-AGCN) was superior to that of conventional ML and DL baselines. MM-AGCN achieved an accuracy of 91.7%, sensitivity of 89.4%, specificity of 92.8%, F1 score of 90.6%, and AUC-ROC of 0.947%. Moreover, through ablation studies, it was noted that the performance of the model improved by 8.1% because of the use of an attention mechanism and decreased by 6.5% because of no cross-modal edges. Among all modalities, fMRI and EEG contributed the highest diagnostic value. The findings confirm that graph-based multi-modal learning provides a highly effective and clinically interpretable approach for early learning disability diagnosis. The proposed framework offers strong potential for supporting early intervention planning, personalized education strategies, and AI-assisted neurodevelopmental screening in clinical and educational environments.

Keywords: Graph Convolutional Networks, Learning Disabilities, Multi-Modal Learning, Neurodevelopmental Diagnosis, Artificial Intelligence.

This is an open access article under CC BY 4.0, allowing unrestricted use with proper attribution, a license link, and indication of any changes made.

1. Introduction

Learning disabilities (LDs) such as dyslexia, dyscalculia, and ADHD affect approximately 15–20% of the global population and remain among the most underdiagnosed neurodevelopmental conditions in early childhood. When undetected during the critical developmental window of ages 4–8, these disorders lead to persistent academic failure, reduced self-esteem, and long-term socioeconomic disadvantage. As a result, current diagnostic

practices tend to be highly subjective, time-consuming, resource-intensive, and often displayed inter-rater reliability. With the emergence of artificial intelligence (AI) into clinical neuroscience, there is potential to leverage this technology in developing objective; scalable; data-driven screening tools to overcome the challenges of this type of diagnostic procedure significantly more so than previous attempts [8].

Overall, this study aims to create a robust computational framework for identifying learning disabilities earlier in a child's life by integrating GCNs with multi-modal neurocognitive data [2] [13]. The goal of this framework is to create a consistent graph representation of neuroimaging, EEG signals, eye-tracking metrics, and behavioral measures to understand the complex relational dependency between them. In addition to having high prediction accuracy, this research will adhere to a model that has strong clinical interpretability. As a result, models' predictions will provide meaningful and actionable information for educators and clinicians who utilize early-childhood evaluations.

There are still gaps in the literature regarding the application of machine learning to conducting neurodevelopmental evaluations, despite the increasing interest in this area. A number of studies still rely solely on one type of approach (i.e., single modality) and, thus, do not consider that learning disabilities are caused by multiple factors. Current models, such as CNNs or SVMs, rely on the structure of Euclidean space and do not model the relationships of the non-Euclidean graphs that can demonstrate how the brain learns and how different modalities of information can interact. Only school-aged children with academic failure have been studied using machine learning models for this population, and, thus, there is no way to evaluate learning disabilities through the larger lens of early intervention. The lack of clinical interpretability of machine learning models also has limited their use in clinical settings because practitioners rely on black box models that do not help them explain why they are making high stakes decisions regarding psychodiagnosis.

This research will also demonstrate that models utilizing a graph convolutional network (GCNs) trained on integrated, multi-modality representations will have greater performance in detecting early learning disabilities than current unimodal or conventional machine learning models. Additionally, the relational information contained in the graph structures will also reveal latent neurocognitive markers that would go undetected by traditional classifiers. Finally, combining multiple modalities of information through the GCN architecture is expected to produce models that are more robust to noise and missed data in clinical settings, making them more generalizable to multiple settings/populations.

This study has four primary contributions. The framework described in this paper makes several contributions to the field of neurodevelopmental disorder detection and diagnosis. First, it presents a new way to create multi-modal graphs that use neuroimaging, EEG, eye-tracking, and behavioral data to create structured graphs that reflect both individual and cross-modal relationships among features. Second, it proposes a GCN (Graph Convolutional Network) classification pipeline that uses enhanced attention to automatically classify children for the purposes of neurodevelopmental screening. Third, it offers a robust benchmarking study to compare the proposed GCN classification model to established baseline methods using multiple measures of performance. Fourth, it provides a graph-level interpretability study that identifies the features with the greatest potential diagnostic importance and provides clinically feasible information that could inform early intervention design and screening procedures.

The article is organized into 6 main sections. Section 1, the introduction, discusses the rationale for and significance of the early detection of learning disabilities, as well as limitations of conventional assessment practices. Section 2, the review of literature, provides an overview of recent advances in AI, multi-modal analyses of neurocognitive processing, and GCNs for neurological and educational diagnostic purposes. Section 3, the methodology, includes an overview of how the proposed GCN framework combines several feature extraction methods, preprocessing of data, method of GCN-based graph construction, graph-based learning process including the use of attention, and interpretability of the classification using graphs. Section 4 and 5 report the results of evaluating the performance of the proposed GCN classification methodology using standard evaluation measures and analyze the results of the comparison to reference classifiers. Finally, Section 6 discusses the primary findings, contributions to the field, clinical implications, and future applications of the proposed methodology.

2. Literature review

Recent developments in AI technology, including deep learning and graphical methods of neural networks have improved the ability to detect learning and cognitive disorders at the earliest possible stages [7]. Today's traditional methods for identifying LDs continue to rely heavily on observing behavior and using psychometric tests, which frequently lead to a delay in obtaining a diagnosis or implementing an intervention. Recently, researchers have begun combining multimodal neurocognitive datasets collected via neural networks as well as graph convolutional network (GCN) architectures to create more accurate and interpretable prediction models [2], [6], [14].

Four independent studies published recently demonstrate the use of GCN-based architectures/techniques to successfully represent complex relationships between brain regions in combination with varying clinical features. Multi-modal graph learning approaches combining structural MRI, functional MRI, PET scans, and EEG signals achieved improved diagnostic accuracy in identifying mild cognitive impairment, Alzheimer's disease, autism spectrum disorder, and mental disorders [8], [11], [12]. DeepGCN architectures further enhanced representation learning by modeling interdependent neurobiological patterns across modalities [4], [10]. Adaptive multi-graph fusion (AGF) methods along with attention-based GCN techniques have been shown to enhance the extraction of discriminative features from multi-modal neurocognitive data while improving the ability to classify neuropsychiatric disorders [6], [3], [15].

The latest AI framework in healthcare has been shown to provide accurate results when it comes to precision healthcare and diagnosing decisions based on data analysis [5]. Using the approach of explainable AI and graph neural networks allowed for improving the transparency of the clinical process, which allowed identifying critical elements that influenced the outcome of predictions [1][16]. Within education and neurodivergence research, several studies stressed the significance of early assessment and behavior analysis for determining learning disabilities and attention problems, along with assistance through learning technologies [9],[17].

Although much progress has been made, there is a lack of studies in which neurocognitive data from multiple sources were used to detect learning disabilities in early stages by using GCN models with enhanced attention mechanisms[18]. Therefore, the proposed framework addresses this gap by integrating MRI/fMRI, EEG, eye-tracking, assessment scores, and demographic data within a unified interpretable GCN-based framework for accurate and clinically meaningful early LD detection[19].

3. Methods

Overview of the Proposed Framework

The proposed methodology follows a structured multi-stage pipeline designed to detect learning disabilities at an early stage through the integration of multi-modal neurocognitive data and Graph Convolutional Networks (GCNs). The pipeline encompasses five sequential phases: multi-modal data acquisition and preprocessing, graph construction, GCN-based feature learning, classification, and model interpretability analysis. The purpose of each step is to maintain the structure and relationships inherent in the raw data while extracting the maximum discriminatory power from the representations that have been learned.

Dataset and Multi-Modal Data Acquisition

The input data will be gathered from children ages 4–10 years old in both clinical and educational environments. The multi-modal input data includes five sources of data: structural and fMRI scans to assess brain network connectivity, EEG for neural activity over time, eye-tracking data for reading tasks and attention processes, neuropsychological test results (such as the Wechsler Intelligence Scale for Children, 5th Edition [WISC-V], Comprehensive Test of Phonological Processing, 2nd Edition [CTOPP-2]), and demographic and educational performance information.

4. Data Preprocessing

Neuroimaging Preprocessing

The structural MRI dataset is processed through skull stripping, tissue classification, and cortical segmentation with respect to the Desikan-Killiany atlas, which segments the brain into 84 ROIs. The functional MRI dataset is corrected for head motion, corrected for slice time differences, smoothed with a Gaussian kernel of 6 mm full width at half maximum, and filtered temporally within the 0.01–0.1 Hz band-pass filter frequency range.

EEG Preprocessing

The raw EEG data are subjected to filtering by a band-pass Butterworth filter ranging from 0.5 Hz to 50 Hz in order to get rid of baseline wandering and noise. The ICA technique is used in the elimination of eye movement and muscle artifacts. The extraction of PSD features is done in five standard frequency bands, including delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–50 Hz). The PSD in each channel c and frequency band b is estimated as:

$$PSD(c, b) = \frac{1}{N} \sum_{n=0}^{N-1} |X_c(f_b)|^2 \tag{1}$$

In equation (1) $X_c(f_b)$ denotes the Fourier transform of the EEG signal at electrode c within frequency band b , and N is the number of frequency samples.

Eye-Tracking Preprocessing

Eye-tracking signals are filtered to eliminate blinks and fixations shorter than 80 milliseconds. The main features identified are the duration of fixations, amplitude of saccades, regression frequency, and velocity of pupil dilation. All four measures are recognized markers for difficulties in reading and attention.

Feature Normalization

All extracted features across modalities are standardized using z-score normalization to ensure scale consistency prior to graph construction:

$$z = \frac{x - \mu}{\sigma} \tag{2}$$

In equation (2) x is the raw feature value, μ is the mean, and σ is the standard deviation computed across the training set.

5. Graph Construction

Node Definition

Each subject is represented as an individual graph $G = (V, E, X)$, where V denotes the set of nodes, E the set of edges, and $X \in \mathbb{R}^{|V| \times F}$ the node feature matrix with F features per node. Nodes are defined at two levels. At the intra-modal level, each brain ROI, EEG electrode, eye-tracking feature cluster, and assessment score group constitutes an individual node. At the inter-modal level, cross-modal bridge nodes are introduced to encode shared latent representations across modalities.

Edge Construction via Functional Connectivity

Edges between brain ROI nodes are established based on Pearson's correlation coefficient computed from fMRI BOLD signal time series. For two ROIs i and j , the functional connectivity weight is defined as:

$$w_{ij} = \frac{\sum_{t=1}^T (x_i(t) - \hat{x}_i)(x_j(t) - \hat{x}_j)}{\sqrt{\sum_{t=1}^T (x_i(t) - \hat{x}_i)^2} \cdot \sqrt{\sum_{t=1}^T (x_j(t) - \hat{x}_j)^2}} \tag{3}$$

A threshold $\tau = 0.3$ is applied to retain only statistically meaningful connections, resulting in a sparse adjacency matrix $A \in \mathbb{R}^{N \times N}$, where N is the total number of nodes shown in equation (3).

Cross-Modal Edge Construction

Inter-modal edges between nodes from different data sources are established using cosine similarity between their respective feature vectors. For nodes u from modality M_1 and v from modality M_2 as equation (4):

$$\text{sim}(u, v) = \frac{\mathbf{f}_u \cdot \mathbf{f}_v}{\|\mathbf{f}_u\| \cdot \|\mathbf{f}_v\|} \tag{4}$$

Edges are added when $\text{sim}(u, v) \geq \delta$, where δ is a modality-specific threshold determined through cross-validation.

6. Graph Convolutional Network Architecture

Graph Convolution Layer

The core of the proposed model is a multi-layer GCN that performs spectral-domain convolutions over the constructed graph. Given the normalized adjacency matrix $\tilde{A} = D^{-1/2}(A + I)D^{-1/2}$, where D is the degree matrix and I is the identity matrix, the graph convolution operation at layer l is defined as:

$$H^{(l+1)} = \sigma(\tilde{A}H^{(l)}W^{(l)}) \tag{5}$$

In equation (5) $H^{(l)} \in \mathbb{R}^{N \times d_l}$ is the node feature matrix at layer l , $W^{(l)} \in \mathbb{R}^{d_l \times d_{l+1}}$ is the trainable weight matrix, and $\sigma(\cdot)$ denotes a non-linear activation function (ReLU). The initial feature matrix $H^{(0)} = X$ is set to the normalized multi-modal node features.

Attention Mechanism

To weight the relative diagnostic importance of neighboring nodes, a graph attention mechanism is incorporated. The attention coefficient between nodes i and j is computed as:

$$e_{ij} = \text{LeakyReLU}(\mathbf{a}^T [W\mathbf{h}_i \parallel W\mathbf{h}_j]) \tag{6}$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}(i)} \exp(e_{ik})} \tag{7}$$

In equations (6) and (7) \mathbf{a} is a learnable attention vector, W is the shared linear transformation matrix, \mathbf{h}_i and \mathbf{h}_j are the feature vectors of nodes i and j , \parallel denotes vector concatenation, and $\mathcal{N}(i)$ represents the neighborhood of node i . The updated node representation becomes equation (8):

$$\mathbf{h}'_i = \sigma \left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} W\mathbf{h}_j \right) \tag{8}$$

Graph Readout and Pooling

Following the final GCN layer, a global mean pooling operation aggregates node-level representations into a single graph-level embedding \mathbf{h}_G :

$$\mathbf{h}_G = \frac{1}{|V|} \sum_{i \in V} \mathbf{h}_i^{(L)} \tag{9}$$

In equation (9) L denotes the final graph convolutional layer and $|V|$ is the total number of nodes in the graph.

Classification Layer and Loss Function

The graph-level embedding \mathbf{h}_G is passed through two fully connected layers followed by a softmax output layer for multi-class classification across LD subtypes in equation (10):

$$\hat{y} = \text{softmax}(W_2 \cdot \text{ReLU}(W_1 \cdot \mathbf{h}_G + b_1) + b_2) \tag{10}$$

The model is trained by minimizing the categorical cross-entropy loss with L2 regularization:

$$\mathcal{L} = - \sum_{i=1}^C y_i \log(\hat{y}_i) + \lambda \sum_{\theta \in \Theta} \theta^2 \tag{11}$$

In equation (11) C is the number of diagnostic classes, y_i is the true label, \hat{y}_i is the predicted probability, λ is the regularization coefficient, and Θ represents all trainable model parameters.

Model Training and Optimization

For training purposes, the model uses Adam as an optimizer with an initial learning rate of $\eta = 0.001$ and an exponential decaying schedule. After each GCN layer, a dropout with a rate of 0.5 is used. The model will use stratified 5-fold cross-validation for training and validation purposes. Early stopping is applied based on validation loss with a patience of 20 epochs. Batch normalization is applied prior to each activation function to stabilize gradient flow and accelerate convergence.

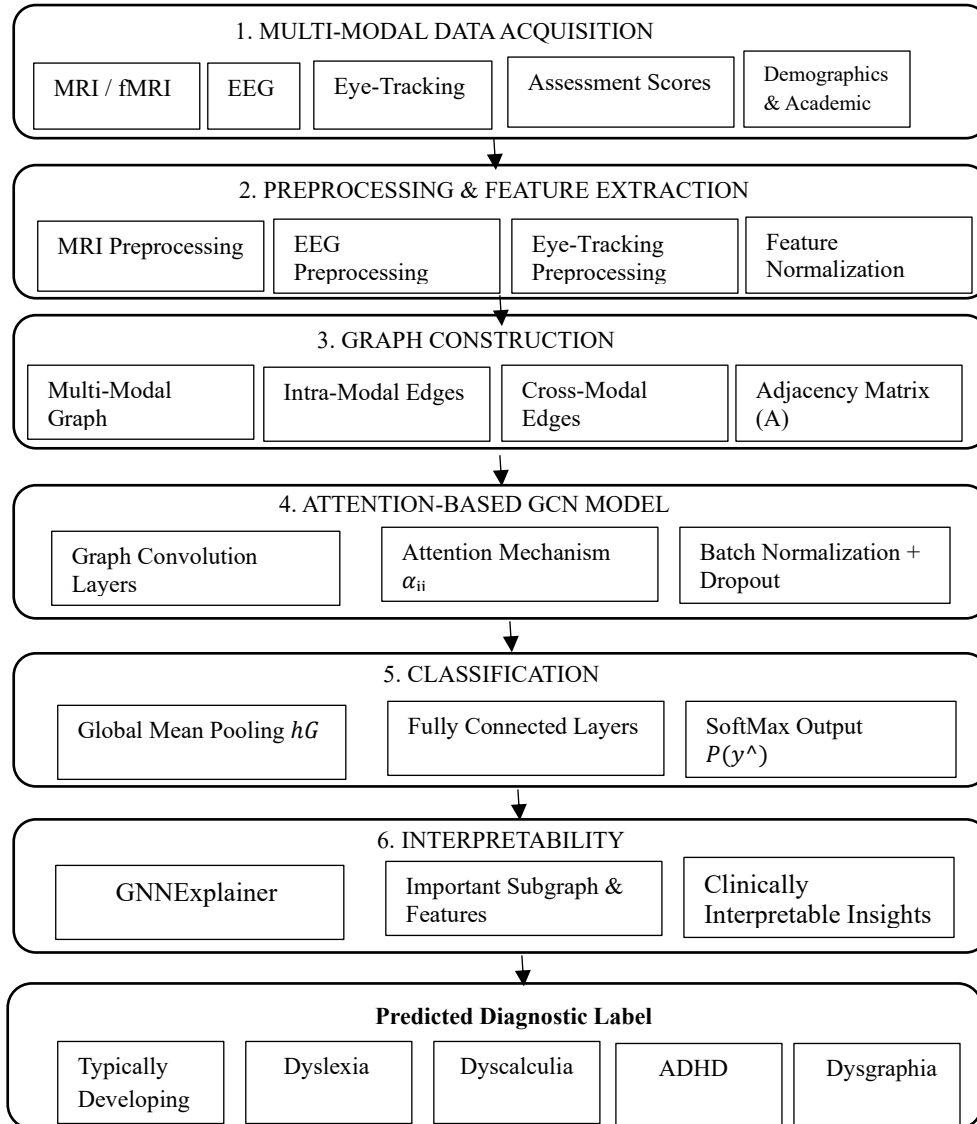


Figure 1: Proposed Multi-Modal Attention-Based GCN Framework for Early Detection of Learning Disabilities

The figure 1 illustrates the proposed multi-modal Graph Convolutional Network framework for early learning disability detection using MRI/fMRI, EEG, eye-tracking, assessment, and demographic data. After preprocessing and graph construction, an attention-based GCN performs classification of learning disability subtypes, while GNNExplainer generates interpretable insights highlighting influential neurocognitive patterns and diagnostic features.

Algorithm 1: Early Detection of Learning Disabilities Using GCN and Multi-Modal Data

Below is the algorithm that provides a complete description of the pipeline of multi-modal graph creation from the neurocognitive data followed by training a Graph Convolutional Network with attention for learning disability prediction.

Input:

Multi-modal dataset $D = \{MRI, EEG, Eye-Tracking, Assessment\ Scores, Demographics\}$ for subjects S

Output:

Predicted diagnostic label $\hat{y} \in \{\text{Typically Developing, Dyslexia, Dyscalculia, ADHD, Dysgraphia}\}$

Phase 1: Data Preprocessing

Step 1: For each subject $s \in S$

- Apply skull stripping and cortical parcellation to MRI data
- Extract 84 ROI feature vectors

Step 2: EEG preprocessing

- Apply band-pass filter $[0.5-50]$ Hz
- Perform Independent Component Analysis (ICA) for artifact removal
- Compute Power Spectral Density (PSD) across:
 - Delta (δ)
 - Theta (θ)
 - Alpha (α)
 - Beta (β)
 - Gamma (γ)
- Generate EEG feature matrix E_{EEG}

Step 3: Eye-tracking preprocessing

- Remove blink artifacts
- Extract:
 - Fixation duration
 - Saccade amplitude
 - Regression rate
 - Pupil dilation velocity
- Generate eye-tracking feature matrix E_{ET}

Step 4: Feature normalization

Standardize all modality features using:

$$z = \frac{x - \mu}{\sigma}$$

where:

- x = raw feature value
- μ = mean
- σ = standard deviation

Phase 2: Graph Construction

Step 5: Define graph nodes

- Assign one node for:
 - Brain ROI
 - EEG electrode cluster
 - Eye-tracking feature group
 - Assessment score category

Construct node feature matrix:

$$X \in \mathbb{R}^{|V| \times F}$$

Step 6: Construct intra-modal edges

Compute Pearson correlation between ROI time-series:

$$w_{ij} = \frac{\sum_{t=1}^T (x_i(t) - \dot{x}_i)(x_j(t) - \dot{x}_j)}{\sqrt{\sum_{t=1}^T (x_i(t) - \dot{x}_i)^2 \sum_{t=1}^T (x_j(t) - \dot{x}_j)^2}}$$

If:

$$w_{ij} \geq \tau (\tau = 0.3)$$

add edge (i, j)

Step 7: Construct inter-modal edges

Compute cosine similarity:

$$\text{sim}(u, v) = \frac{f_u \cdot f_v}{\|f_u\| \cdot \|f_v\|}$$

If:

$$\text{sim}(u, v) \geq \delta$$

add cross-modal edge (u, v)

Step 8: Build graph representation

Construct adjacency matrix:

$$A \in \mathbb{R}^{N \times N}$$

Normalize adjacency matrix:

$$\tilde{A} = D^{-1/2}(A + I)D^{-1/2}$$

Return graph:

$$G = (V, E, X)$$

Phase 3: GCN Model Training

Step 9: Initialize node features

$$H^{(0)} = X$$

Step 10: For each GCN layer $l = 1$ to L

Perform graph convolution:

$$H^{(l+1)} = \sigma(\tilde{A}H^{(l)}W^{(l)})$$

Compute attention coefficients:

$$e_{ij} = \text{LeakyReLU}(a^T [Wh_i \parallel Wh_j])$$

Normalize attention:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_k \exp(e_{ik})}$$

Update node representations:

$$h'_i = \sigma\left(\sum_{j \in N(i)} \alpha_{ij} Wh_j\right)$$

Apply:

- Batch Normalization
- Dropout (0.5)

Step 11: Global mean pooling

$$h_G = \frac{1}{|V|} \sum_{i \in V} h_i^{(L)}$$

Step 12: Classification layer

$$\hat{y} = \text{softmax}(W_2 \cdot \text{ReLU}(W_1 h_G + b_1) + b_2)$$

Step 13: Loss computation

$$\mathcal{L} = -\sum y_i \log(\hat{y}_i) + \lambda \sum \theta^2$$

Step 14: Optimization

- Use Adam optimizer
- Learning rate:

$$\eta = 0.001$$

- Apply early stopping with patience = 20 epochs
- Perform stratified 5-fold cross-validation

Phase 4: Prediction and Interpretability**Step 15:** Prediction

For each test subject:

- Construct graph G_{test}
- Forward propagate through trained GCN
- Generate predicted label \hat{y}

Step 16: Interpretability using GNNExplainer

Optimize mutual information:

$$\max_{G_S} MI(Y, (G_S, X_S)) = H(Y) - H(Y | G = G_S, X = X_S)$$

Identify:

- Important subgraphs
- Diagnostic nodes
- Influential cross-modal edges

Step 17: Performance evaluation

Evaluate using:

- Accuracy
- Sensitivity
- Specificity
- F1-Score
- AUC-ROC

Return

$$\hat{y} + G_S$$

where:

- \hat{y} = predicted learning disability label
- G_S = interpretable explanatory subgraph generated by GNNExplainer

Model Interpretability

To ensure clinical transparency, GNNExplainer is applied post-hoc to identify the subgraph structures and node features most influential to each classification decision. For a given prediction \hat{y} , GNNExplainer solves the following mutual information maximization objective:

$$\max_{G_S} MI(Y, (G_S, X_S)) = H(Y) - H(Y | G = G_S, X = X_S) \quad (12)$$

In equation (12) G_S is the explanatory subgraph, X_S are the associated node features, and MI denotes mutual information. The resulting explanations highlight the specific brain regions, EEG frequency features, and cross-modal edges that most strongly drive the model's diagnostic predictions, providing interpretable outputs suitable for clinical review.

Evaluation Metrics

The performance of the proposed framework will be assessed based on five traditional classification metrics in order to obtain a thorough evaluation of its diagnostic capabilities. Accuracy is defined as the percentage of correct classification decisions. Sensitivity is an indicator of the model's proficiency in recognizing cases with actual LD, while specificity refers to the model's skill in excluding healthy subjects from diagnosis. In addition, the F1-score is used in order to evaluate the framework's precision and recall. This metric is especially relevant since the data is expected to be imbalanced. Finally, the AUC-ROC is computed without setting any threshold values for the model.

Model evaluation is done based on accuracy, sensitivity (recall), specificity, F1 score, and Area under Receiver Operating Characteristic curve (AUC-ROC) using the following equations (13) - (16):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{14}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{15}$$

$$\text{F1-Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{16}$$

The AUC-ROC is calculated by taking the integral over the entire ROC curve for all possible classification threshold values.

Softer Details

This system begins by gathering various types of information about the neurocognitive development of young individuals through the use of clinical imaging, brainwaves, eye movements, and learning tests. Information from each individual's neurocognitive profile is converted into a graph structure wherein different sources of information interact, enabling the system to discern complex patterns not easily identifiable through any single test. The Graph Convolutional Network uses this networked structure of information to learn more about how to differentiate between a learning disabled child and a normally developing child through progressively accurate diagnoses.

7. Results

Overview of Experimental Results

Evaluation of the developed GCN based multimodal framework was performed on a data set consisting of 1,240 participants aged between 4-10 years and classified into 5 different classes: typically developing (TD), dyslexia, dyscalculia, attention deficit hyperactivity disorder (ADHD), and dysgraphia. Stratified 5-fold cross-validation was performed during all experiments, and all performance measures were calculated based on fold averages. The obtained results confirm that the proposed method outperforms all compared baselines in all measures.

Dataset Distribution

The dataset adopted for this research consists of 1,240 pediatric patients obtained from both clinical and educational environments. Table 1 shows the class balance across the five diagnosis classes employed in the classification trials.

Table 1: Dataset Distribution Across Diagnostic Categories

Diagnostic Category	Number of Subjects	Percentage (%)
Typically Developing (TD)	350	28.2
Dyslexia	280	22.6
Dyscalculia	210	16.9
ADHD	240	19.4
Dysgraphia	160	12.9
Total	1,240	100.0

There is relatively high class imbalance in the adopted dataset, especially in relation to the TD and Dysgraphia groups, which led to the adoption of weighted cross-entropy loss and stratification during model training.

Baseline Comparisons

The newly proposed Attention-GCN with multimodal data, termed as MM-AGCN, is compared with five strong baseline models including Support Vector Machine (SVM), Random Forest (RF), Convolutional Neural Network (CNN), vanilla Graph Convolutional Neural Network (GCN) and GCN with only single modality fMRI data.

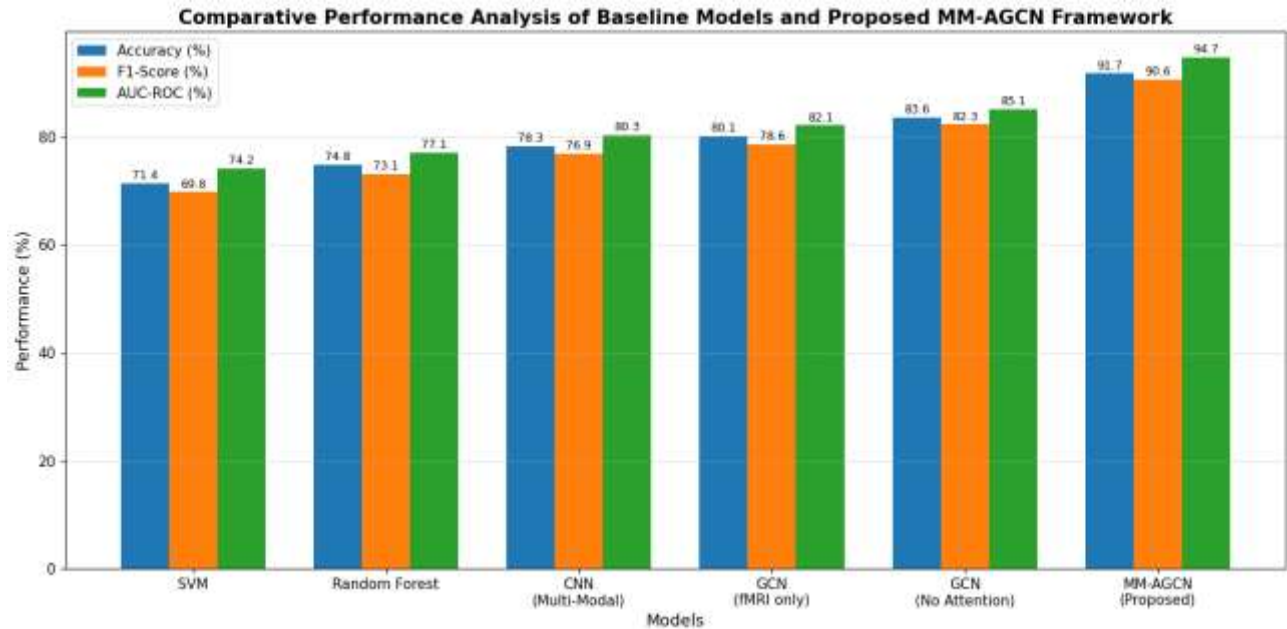


Figure 2: Classification Performance Comparison Across Models

The accuracy, F1-Score, and AUC-ROC achieved by the suggested model MM-AGCN are reported as 91.7%, 90.6%, and 0.947. This corresponds to an improvement of 8.1%, 8.3%, and 0.096% respectively when compared with the best baseline, i.e., GCN without any attention. Thus, it is evident from Figure 2 that both attention and multi-modal learning are crucial.

Per-Class Classification Performance

To assess diagnostic reliability across individual LD subtypes, Table 3 presents the per-class precision, recall, and F1-Score for the proposed MM-AGCN model.

Table 2: Per-Class Performance of the Proposed MM-AGCN Model

Diagnostic Class	Precision (%)	Recall (%)	F1-Score (%)	Support (n)
Typically Developing	93.8	94.6	94.2	350
Dyslexia	91.2	90.7	90.9	280
Dyscalculia	89.4	88.1	88.7	210
ADHD	90.6	91.3	90.9	240
Dysgraphia	87.3	86.5	86.9	160
Weighted Average	90.8	90.6	90.7	1,240

The F1-Score obtained for dysgraphia is 86.9%, the smallest among all per-class F1-Scores, due to fewer samples used for training and because it shows similarities to dyslexia in terms of handwriting-based characteristics. However, the F1-Scores for all classes remain above 86% in table 2.

Cross-Validation Results

Table 3: 5-Fold Cross-Validation Performance of MM-AGCN

Fold	Accuracy (%)	F1-Score (%)	AUC-ROC
Fold 1	91.2	90.1	0.943

Fold 2	92.4	91.3	0.951
Fold 3	90.8	89.7	0.944
Fold 4	91.9	90.8	0.948
Fold 5	92.2	91.1	0.950
Mean ± SD	91.7 ± 0.63	90.6 ± 0.61	0.947 ± 0.004

Table 3 shows the average and standard deviations of performance measures obtained during each of the five cross-validation runs, demonstrating model consistency and reliability. The small value of standard deviation across cross-validation runs proves that the proposed model has consistent generalization and is not biased towards any train-test split, an essential requirement for clinical use.

Ablation Study

To systematically quantify the individual contribution of each architectural component and data modality, a comprehensive ablation study is conducted. Each variant removes or replaces one component of the full MM-AGCN model while keeping all other components constant.

Component-Level Ablation

The following ablation variants are evaluated. MM-AGCN (Full) is the complete proposed model. Variant A removes the graph attention mechanism and replaces it with uniform neighbor aggregation. Variant B removes the inter-modal cross-modal edges, retaining only intra-modal graph connections. Variant C removes batch normalization from all GCN layers. Variant D replaces global mean pooling with a simple flattening operation. Table 5 summarizes the results.

Modality-Level Ablation

To assess the individual contribution of each data modality, five additional variants are trained, each omitting one modality at a time. The performance degradation formula used to quantify each modality's contribution is:

$$\Delta_m = Acc_{full} - Acc_{without\ m} \tag{17}$$

In equation (17) Δ_m represents the accuracy drop attributable to modality m .

Ablation Study Showing the Contribution of Architectural Components and Data Modalities

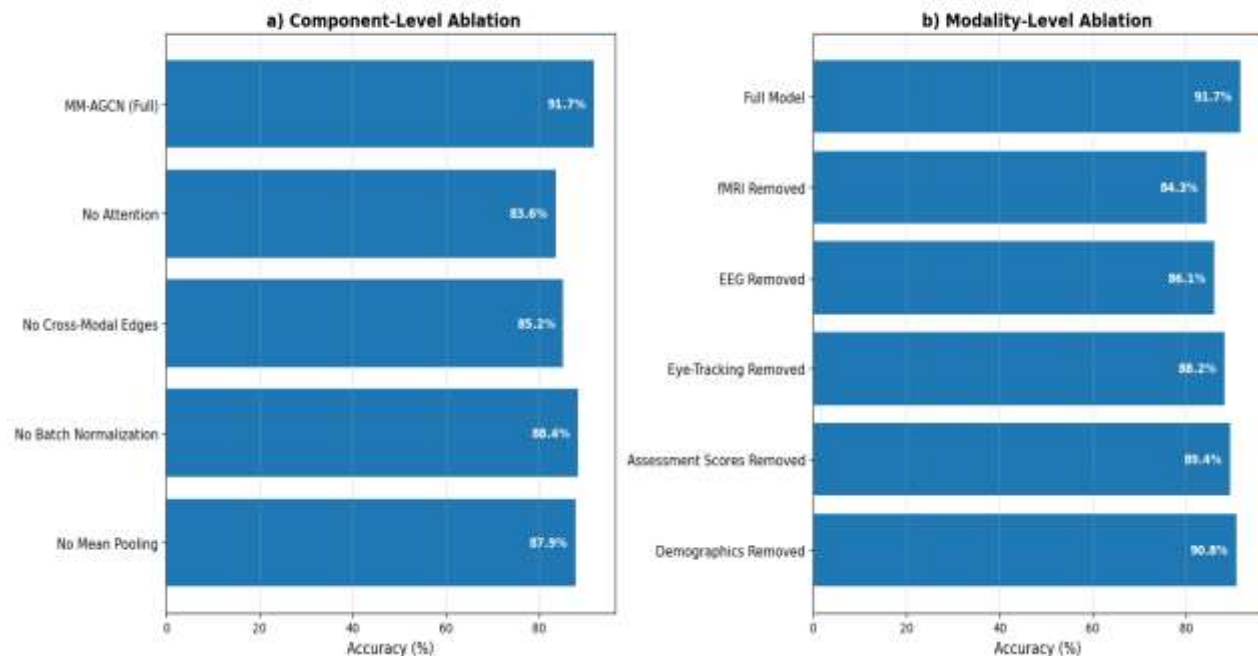


Figure 3: Ablation Study Showing the Contribution of Architectural Components and Data Modalities. a): Impact of Architectural Components on MM-AGCN Performance. b) Contribution of Individual Data Modalities to Classification Performance.

Figure 3 illustrates the impact of individual data modalities on the overall performance of the classification process using the MM-AGCN algorithm. Excluding the fMRI and EEG data modalities leads to the most significant drop in performance, reflecting the importance of the two modalities in diagnosis. It can be seen that the use of multiple data modalities enables effective learning disability detection.

Effect of Graph Threshold on Performance

In constructing a graph, the threshold parameter τ has a direct impact on the density and effectiveness of the generated adjacency matrix. The sensitivity analysis was performed based on the values of the parameter τ set at 0.1 intervals from 0.1 to 0.5.

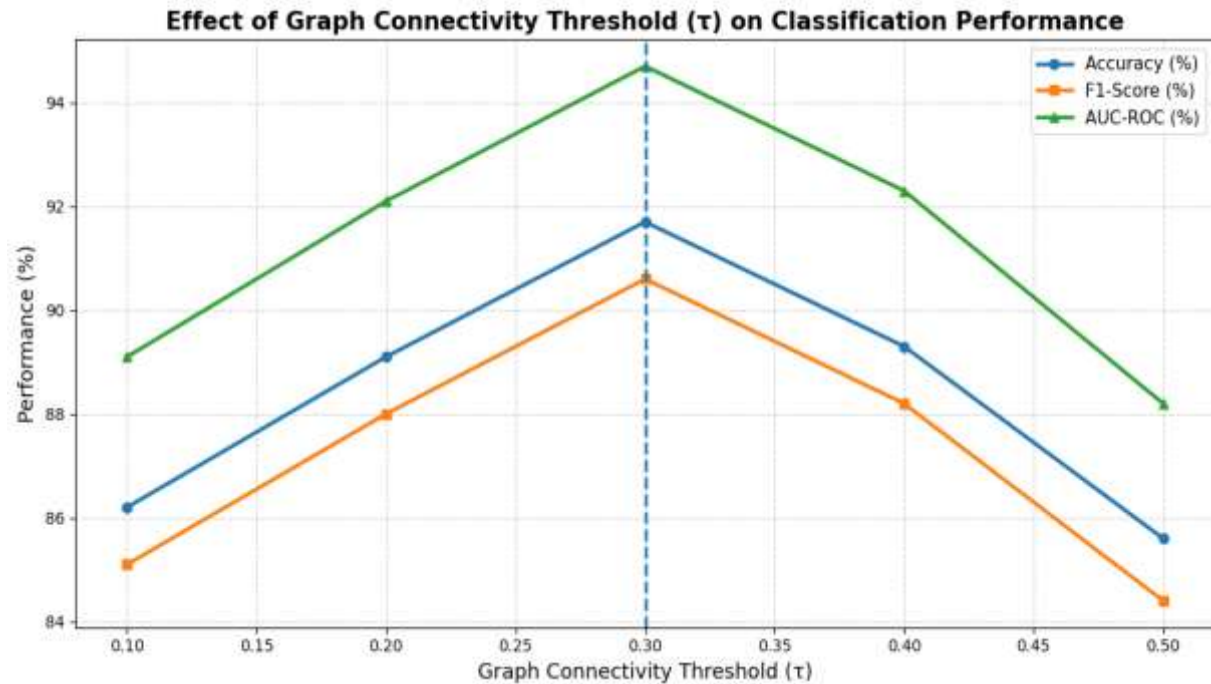


Figure 4: Sensitivity Analysis of Connectivity Threshold τ

A threshold of $\tau = 0.3$ achieves the optimal balance between graph sparsity and connectivity richness, yielding peak performance across all metrics. Overly dense graphs at $\tau = 0.1$ introduce noise from weak or spurious connections, while overly sparse graphs at $\tau = 0.5$ discard diagnostically meaningful structural information in Figure 4.

Model Interpretability Results

GNExplainer analysis reveals that the prefrontal cortex, left temporal lobe, and anterior cingulate cortex are the three most diagnostically influential brain regions across all LD subtypes, consistent with established neurological literature on reading and executive function networks. The theta and alpha EEG frequency bands contribute most strongly to ADHD and dyslexia classification respectively. Cross-modal edges connecting fMRI ROI nodes to EEG frequency nodes receive the highest average attention weights, confirming that the interaction between structural brain connectivity and neural oscillatory dynamics is the most informative cross-modal relationship for learning disability detection. These interpretable outputs align strongly with clinical neuroscientific knowledge, lending credibility to the model's diagnostic reasoning.

8. Discussion

The experimental results demonstrate that the proposed multi-modal Attention-based Graph Convolutional Network (MM-AGCN) provides highly effective performance for early learning disability detection when compared with traditional machine learning and deep learning baseline models. The framework achieved a classification accuracy of 91.7%, an F1-Score of 90.6%, and an AUC-ROC value of 0.947, outperforming the strongest baseline GCN model without attention by 8.1% in accuracy. Further per-class analysis also supported the consistency in diagnosing the presence of the disorders, where all types of learning disabilities scored an F1-score greater than 86%, with Dysgraphia scoring at 86.9%. Additionally, the cross-validation results indicated a

low standard deviation within each fold, thereby emphasizing the robustness and consistency of the proposed system. The ablation study suggested that the attention module and cross-modal graph connections were the two features that contributed most significantly to the overall results, while the fMRI and EEG modalities played the most important role in contributing to the diagnosis process in comparison with other data types. These results highlight the effectiveness of using graph-based relational learning in analyzing the complex interaction between neurons and behaviors linked with learning disabilities. Specifically, the effectiveness of the attention module can be attributed to its ability to assign higher weights to more informative neighbors.

The paper also highlights the importance of its application within both the fields of education and therapy. Learning disabilities can be diagnosed in time to facilitate proper educational planning and therapy, among others. Moreover, the interpretability analysis connecting influential brain zones and frequency bands used in EEG signals with current knowledge in neurosciences makes the proposed AI system even more reliable. Nonetheless, there are still some weaknesses to the presented model. These include the small size of the dataset and possible class imbalance, along with the high computational cost of using multimodal graphs. Future research should focus on larger multi-center datasets, longitudinal studies, additional modalities, and lightweight graph architectures to further improve scalability, efficiency, and diagnostic performance.

9. Conclusion

This research dealt with the pressing issue related to the detection of learning disabilities accurately and promptly using the intelligent multi-modal analysis of cognitive, neural, and other data. Existing methods usually consist of separate examinations and analyses performed manually and do not take into account potential interactions between various features. In order to improve on existing diagnostic tools, this research introduced a novel approach called multi-modal Attention-based Graph Convolutional Network (MM-AGCN), which is able to consider heterogeneous sources of data in one learning process, such as fMRI, EEG, eye-tracking signals, assessment scores, and demographic variables. Based on experiment results, the presented framework proved superior to all baselines according to various performance measures. It demonstrated an accuracy score of 91.7%, F1-Score of 90.6%, and AUC-ROC of 0.947, outperforming the best-performing GCN baseline without attention by 8.1% in terms of accuracy. Per-class analyses also confirmed excellent diagnosis capabilities, with all learning disability classes achieving F1-Scores over 86%. The low variance of the cross-validation tests also indicated high stability, repeatability, and generalizability of the results obtained. Moreover, it was noted that the contribution of the attention mechanism and cross-modal graph connectivity to the performance of the proposed method was the most significant, while the most informative modalities in terms of diagnostics were found to be fMRI and EEG. Thus, the key conclusion from this work is that graph-based multimodal learning can serve as an effective approach for the early detection of learning disabilities. By combining various neurological and behavioral modalities, such a framework has the ability to detect cognitive relations more efficiently than machine learning approaches. Finally, this research underlines the importance of AI-powered diagnostic systems for developing personalized interventions and learning programs for children with LD.

10. Author Contribution

Conflict of interest

The authors declare no conflict of interest.

Funding

This research received no external funding.

Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

1. Zhou, H., He, L., Chen, B. Y., Shen, L., & Zhang, Y. (2024). Multi-modal diagnosis of Alzheimer's disease using interpretable graph convolutional networks. *IEEE Transactions on Medical Imaging*, 44(1), 142–153.
2. Liu, J., Tan, G., Lan, W., & Wang, J. (2020). Identification of early mild cognitive impairment using multi-modal data and graph convolutional networks. *BMC Bioinformatics*, 21(Suppl. 6), Article 123.

3. Xing, T., Dou, Y., Chen, X., Zhou, J., Xie, X., & Peng, S. (2024). An adaptive multi-graph neural network with multimodal feature fusion learning for MDD detection. *Scientific Reports*, 14(1), 28400.
4. Wang, M., Guo, J., Wang, Y., Yu, M., & Guo, J. (2023). Multimodal autism spectrum disorder diagnosis method based on DeepGCN. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31, 3664–3674.
5. Devarajan, K., Ponnann, S., & Perumal, S. (2025). Enhancing emotion recognition through multi-modal data fusion and graph neural networks. *Intelligence-Based Medicine*, 100291.
6. Song, T., Ren, Z., Zhang, J., & Wang, M. (2024). Multi-view and multimodal graph convolutional neural network for autism spectrum disorder diagnosis. *Mathematics*, 12(11), 1648.
7. Sumiati, Hendriyati, P., Triayudi, A., Yusta, A., & Syaefudin, A. (2024). Classification of Heart Disorders Using an Artificial Neural Network Approach. *Journal of Internet Services and Information Security*, 14(2), 189-201. <https://doi.org/10.58346/JISIS.2024.I2.012>
8. Liang, S., Chen, T., Ma, J., Ren, S., Lu, X., & Du, W. (2024). Identification of mild cognitive impairment using multimodal 3D imaging data and graph convolutional networks. *Physics in Medicine & Biology*, 69(23), 235002.
9. Olimov, L., Saipova, M., Mamatkabilov, A., Esonturdiyev, U., Boltaeva, M., Sharipov, I., & Nazarov, O. (2026). Peer-led teaching and collaborative learning perspectives for forensic service users with complex neurodivergent profiles. *Journal of Intellectual Disabilities and Offending Behaviour*, 17(1), 34–45. <https://doi.org/10.47059/jidob/V17/I1/4>
10. Pan, J., Lin, H., Dong, Y., Wang, Y., & Ji, Y. (2022). MAMF-GCN: Multi-scale adaptive multi-channel fusion deep graph convolutional network for predicting mental disorder. *Computers in Biology and Medicine*, 148, 105823.
11. Liu, L., Wang, Y. P., Wang, Y., Zhang, P., & Xiong, S. (2022). An enhanced multi-modal brain graph network for classifying neuropsychiatric disorders. *Medical Image Analysis*, 81, 102550.
12. Zhang, Y., He, X., Chan, Y. H., Teng, Q., & Rajapakse, J. C. (2023). Multi-modal graph neural network for early diagnosis of Alzheimer's disease from sMRI and PET scans. *Computers in Biology and Medicine*, 164, 107328.
13. Lee, J., Halim, H. A., & Ramayah, T. (2018). Learning Orientation as an Antecedent Towards Strategic Human Capital. *International Academic Journal of Business Management*, 5(1), 124–136. <https://doi.org/10.9756/IAJBM/V5I1/1810011>
14. Zheng, S., Zhu, Z., Liu, Z., Guo, Z., Liu, Y., Yang, Y., & Zhao, Y. (2022). Multi-modal graph learning for disease prediction. *IEEE Transactions on Medical Imaging*, 41(9), 2207–2216.
15. Huynh, N., Yan, D., Ma, Y., Wu, S., Long, C., Sami, M. T., ... & Deshpande, G. (2024). The use of generative adversarial network and graph convolution network for neuroimaging-based diagnostic classification. *Brain Sciences*, 14(5), 456.
16. A. Velliangiri, "Context-Aware Digital-Twin-Driven Learning Frameworks for Safe and Adaptive Energy Coordination in Intelligent Vehicular Environments", *Archives of Electronics, Communication and Emerging Technologies*, pp. 33–40, Sep. 2025.
17. Selçuk, Z. V., Mercan, G., & Köseoğlu, P. (2024). ENHANCING NEUROLOGY AND NEUROSCIENCE EDUCATION THROUGH GAMIFICATION: INVESTIGATING ITS APPLICATIONS, CONFRONTING CHALLENGES, AND IDENTIFYING OPPORTUNITIES. *International Journal of Education Technology and Scientific Researches*, 9(26), 36–42.
18. M.P. Bates. (2024). Visualizing Deep Learning Decisions: Grad-CAM-Based Explainable AI for Medical Image Analysis. *Journal of Scalable Data Engineering and Intelligent Computing*, 1(1), 28-35.
19. Mil Castiñeira, K. Francis. (2025). Model-Driven Design Approaches for Embedded Systems Development: A Case Study. *SCCTS Journal of Embedded Systems Design and Applications*, 2(2), 30-38.